DOCUMENT RESUME

ED 064 316 TM 001 481

AUTHOR Lynch, Mervin D.
TITLE Multidimensional Measurement with the D Statistic and the Semantic Differential.
PUB DATE Apr 72
NOTE 18p.; Paper presented at the annual meeting of the American Educational Research Association (Chicago, Illinois, April 1972)

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Analytic Geometry; Comparative Analysis; Educational Research; *Factor Analysis; *Mathematical Linguistics; *Measurement Techniques; *Statistics
IDENTIFIERS D Statistic; *Semantic Differential

ABSTRACT

Three approaches used in the analysis of semantic differential data to determine similarity in meaning between profile ratings on concept pairs are discussed. These approaches are: (1) to compute mean scores on each scale for each concept and to compare concepts on a scale-by-scale basis; (2) to form mean scores on each dimension for each concept and to compare concepts on a dimensional basis; and (3) to compute the D statistic between ratings on two concepts and use this as a multidimensional measure for comparison purposes. A figure and tables illustrate the three approaches. (DB)

MULTIDIMENSIONAL MEASUREMENT WITH THE D STATISTIC AND THE SEMANTIC DIFFERENTIAL

By

Mervin D. Lynch

In their haste to analyze semantic differential data, researchers have often ignored the multidimensional nature of their measurement technique. They have obtained ratings on semantic differential scales on two or more concepts and have made comparisons between concepts on a scale-by-scale or a dimensions by dimension basis but not on a single index of meaning. As a result, analyses on semantic differential data have been difficult to interpret and of limited theoretical or pragmatic utility.

Three main approaches have been applied to analysis of semantic differential data to determine similarity in meaning between profile ratings on concept pairs. These are 1) to compute mean scores on each scale for each concept and to compare concepts on a scale-by-scale basis; 2) to form mean scores on each dimension for each concept and to compare concepts on a dimensional basis; and 3) to compute the D statistic between ratings on two concepts and use this as a multidimensional measure for comparison purposes. The first two approaches have been widely utilized in semantic differential research and were presented as major approaches for semantic differential data analysis (Osgood et. al., 1951; Osgood and Snider, 1969). The D statistic was proposed as a possible analysis technique

in The Measurement of Meaning, (Osgood, 1957) but this was apparently not a serious suggestion, since the D statistic was not developed or elaborated on in that or subsequent reference works on the semantic differential (Osgood and Snider, 1969).

Scale-by-Scale Analysis

One approach in comparing profiles on semantic differential data has been to compute mean scores on each adjective pair scale for each of two concepts and to compare between concepts with the scale means on a scale-by-scale basis. These mean scores are presented as profiles which are used to make scale wise comparisons between the two concepts. A limited example is included to illustrate the computational techniques and modes of presentation of the different analysis procedures. The numbers in the figures and tables are only illustrative and not data reported as research results. Data on two concepts Myself and My Teachers, on each of seven adjective bi-polar scales for two boys and two girls are shown in Table 1. In this example mean scores for each concept and each scale are computed for the boy and girl sample and these mean scores are presented as profiles in Figure 1.

--Table 1 and Figure 1 About Here--

The temptation in this approach is to interpret the ratings in terms of common sense meaning of each adjective pair, and such common sense meanings may not be the particularistic meanings which were evoked when a rating was made on each scale. The choice of scales for use in semantic differential measures is generally based

on prior factor analytic research and scales so selected are representors of the factors extracted in that factor analysis. For instance, warm-cool and heavy-light are generally representors of the potency dimension of meaning, and fast-slow and active-passive are usually representors of the activity dimension of meaning. Each scale is not uniquely meaningful in its own right although it may sound as if it is and it should be interpreted as one index of the dimension of meaning it represents. For instance, to say as in Figure 1, that the two boys rate teachers as faster than themselves because they had a higher mean score on the scale, slow-fast, does not imply that teachers could beat them in a foot race or in a solution of a computational problem or are socially more adept; rather it means that teachers had a higher score on the activity dimension of meanings according to ratings of these students and that is all.

Dimension by Dimension Analysis

A second approach which has been used to make comparisons on ratings between two concepts with the semantic differential has been to compare between concepts on the basis of mean scores for each dimension. In this approach, mean scores for each factor of meaning are formed by summing across all adjective pairs which are representors of each factor. The factors and scales designative of those factors are determined by means of factor analysis on semantic differential ratings. Tables 2 and 3 present the mean scores by dimension for student ratings on Myself and My Teachers. Scores on evaluation are obtained by adding scores for ratings on scales such

as valuable-worthless, good-bad, and favorable-unfavorable for each person for each concept. Similarly, scores on potency are obtained by adding scores for ratings on scales such as weak-strong and hard-soft.

Separate analysis need to be made for each dimension to compare ratings on the two concepts and these are shown for the example in Table 3. To get at an overall index of similarity in meaning requires an inferential jump for the scores for each of the three factors, but this is not justified in the methodology. Since the activity and potency factors have no theoretical utility, it is possible to only make a statement about the relative attitudes of boys and girls as shown by scores on evaluation for the two concepts.

--Tables 2 and 3 About Here--

Dimensions of meaning have been extracted and named on the basis of factor analysis on ratings obtained with the semantic differential. Three more or less general dimensions of meaning have been identified in prior factor analytic research, (Osgood, 1957; Osgood and Snider, 1969), evaluation, activity, and potency, as well as some relatively unique factors such as personalism, receptivity, and complexity. The names of the dimensions isolated from these factor analyses are labels for hypothetical constructs of meaning which may or may not have theoretical utility. The evaluative dimension of meaning has a direct theoretical utility insofar as ratings on the evaluative dimension may be used as a generalized index of attitudes towards various stimulus objects.

The potency and activity dimensions unlike evaluation have no identified theoretical construct with which they are associated but are useful names for tagging dimensions of meaning. Each of the various dimensions of meaning represent individual aspects of meaning and taken separately they don't serve the same function as the variable of meaning itself. To use data from diverse dimensions such as evaluation, potency and activity for purposes of measuring meaning it is necessary to employ an index that will combine data from the various dimensions in one single multidimensional measure. Dimensions of meaning are like factors of height and weight which cannot provide an overall index of meaning by summing across dimension scores on a concept; rather a multidimensional measure is needed for this purpose and the D statistic provided one such multidimensional measure.

The D Statistic

A preferable approach was to compare meanings of two or more concepts obtained with semantic differential ratings with a multi-dimensional index which combined ratings across scales and dimensions to form a measure with degrees along a concinuum of connotative judgment. The D statistic, the generalized distance function in analytic geometry provides one such measure which may be used for adding over different dimension of meaning. This index was proposed as a multidimensional measure for indexing similarity in meaning by Osgood et. al. (1957) and later has seen more or less limited applications (Tannenbaum and Lynch, 1958; Lazowick, 1955; Lynch et. al. 1968 and Tannenbaum and McLeod, 1967). The D statistic

provides a measure of similarity in meaning between two profiles of judgment such as ratings on two concepts, on a person and a standard or between two persons according to the following formula:

$$D = \sqrt{\sum_{i=1}^{N} (X_{ij} - X_{ik})^2}$$

where (i) is a scale, (j) is a concept, (k) is a concept and (N) is the number of scales.

The D statistic is computed by subtracting the scores for ratings on each scale for two concepts, squaring the resulting differences and summing the squares of the differences. The square root of this sum of squares provides the D statistic which is an index of similarity in meaning between two concepts. For instance, the method of computing D was applied to the semantic differential ratings on the concept Myself and My Teachers as shown in Table 4. In this analysis the scores for each student on each scale for the concept My Teacher are subtracted from the scores on those scales for the concept Myself. The resulting difference is squared and a D value is computed as the square root of the sum of squares of these differences for each of the two boys and two girls. These D values provide an index of similarity in meaning for each person and this index for these two concepts measures teacher-student identification. In this instance, girls apparently identify more with their teachers than boys but the significance of the difference in mean scores remains to be subjected to statistical test with either the t test or analysis of variance.

The D statistic provided measurement on an interval scale and its directionality is interpreted on a relative basis such that two or more concepts are more or less similar in meaning to some third concept. The D statistic is a reciprocal measure in that the smaller the D values the more similar in meaning are two concepts adjudged. For measures such as identification or idealization, smaller D values are interpreted as greater idealization or identification. In the example in Table 4 a smaller D value for girls is interpreted as a greater degree of teacher-student identification. The larger the D value the less similar in meaning are two concepts judged and in the instances of identification and idealization the less the degrees of identification and idealization.

--Table 4 About Here--

Mahalinobis (1937, 1948) has studied a more general form of the D statistic and has shown that the D statistic follows a normal process and has the parameters of the normal distribution, but the distributional nature for semantic differential data needs to be determined. Until a Monte Carlo simulation is made to determine the exact distributional nature of D when used with semantic differential data, it is perhaps better to assume that the D statistic is normally distributed, a likely possibility in most instances, and get on with the research work, than to be mired by the unlikely contingency that the data is so skewed that it may not be analyzed by parametric procedures. In general, parametric statistics such as analyses of variance are sufficiently powerful to accommodate considerable departure from normality should it occur with the use of D.

A major problem in the use of D has been to determine how many and what scales to include in computing the D statistic. To date, scales have been included in terms of the relative proportion of the contribution of each factor to the total common factor variance. The common factor variance is determined by applying factor analysis to the semantic differential ratings each time data is obtained. For instance, if the percent of variance was respectively 40 percent for evaluation, 20 percent for potency and 15 percent for activity, scales would be included in the D on a 4:2:2 basis. Individual scales would be added to D to represent additional common or unique factors. The scales selected for inclusion are those with the highest factor loadings on each factor and relatively high reliabilities. This selection apparently provides scales for D computations which are specifically relevant and representative of the concepts rated.

Some Applications

Teacher student learning activities among other educational processes involve the sharing of meanings vis a vis communication of connotative and denotative meanings between teachers and students. Learning of denotative or factual information aspect of meaning is often the major focus of educational efforts, but learning of these meanings may be mediated by the connotative or feeling tone aspects of meaning. The D statistic when used with the semantic differential provides one way of indexing mediating variables, such as identification idealization, empathy, and others which may be critical in the success or failure of students in learning activities. To date, the D statistic has not been extensively applied, and much of its

application is currently being submitted for publication. There are however, a host of uses to which the D statistic may be put.

A major use of the D statistic has been to provide an index of identification. Lazowick (1955) proposed that the D statistic computed between subject and model for any given concept will provide a useful operationalization for identification. The D value between ratings on concepts Myself and My Parents, Myself and My Teachers, and Myself and Children I Know, provide measures respectively of parent identification, teacher identification, and peer group identification. Bushinski (1969) applied the D statistic to compare identification of students with teachers in student and teacher centered environments and found no appreciable effects. Lynch (1971) applied the D statistic to study identification patterns of creative adolescents with parents, teachers and peers and found an appreciable impact of peer group identification on high as opposed to low creative persons. Nugent (1970) used the D statistic to study factors which mediated student judgments of teacher effectiveness, and found that teacher student identification has an appreciable effect upon these judgments as a mediating variable.

Measures of idealization based on the use of the D statistic have been widely utilized in political election studies (Tannenbaum and Greenberg, 1961; Lynch, 1967) to index candidate image where comparisons were made between a candidate for an office and the ideal candidate for that office. The D statistic also has considerable promise in educational research. The D between ratings on concepts Myself and Myself as I'd Like to Be, My Teacher and The Ideal Teacher

and Myself and The Ideal Administrator provide measures respectively of self idealization, teacher idealization and administrator idealization. Nugent (1970) found that both teacher idealization and identification mediated judgments of teacher effectiveness. Lynch and Kaufman (1971) found that an experimental remedial reading class showed appreciably greater teacher idealization after an 8 weeks remedial experience, and it may be that this idealization mediated their improved performance on reading tasks, over that period of time. Lynch and Lemkuhl (1971) compared the images of public and vocational school administrators in terms of the idealization of their roles and found that the vocational school administrators viewed their role as more similar in meaning to the ideal administrator than did the public school administrators, but that this varied with age, and the older vocational administrators saw their position as more similar in meaning to the ideal.

Other studies have focused on the use of the D statistic as a measure of empathy. In particular D has been applied with the semantic differential to index empathy between parents and children, teachers and students, and children and peers. In this studies the D statistic between Myself and How My Parents See Me, Myself and How My Teachers See Me, and Myself and How Other Children See Me provide measures respectively of parent empathy, teacher empathy, and peer group empathy. Peer group and parent empathy were shown to have a marked impact on the ratings of creative as opposed to less creative adolescents in a study by Lynch (1971).

Studies of stylistic language variables that have particular relevance to research in counselor education, reading and speech

pathology, have utilized the D statistic. In this research ratings were obtained on a passage and a standard, and the D statistic is computed between the two. One such application, Lynch et. al., (1967) applied D as a measure of human interest in writing. A set of scales was selected and successively refined through factor analysis to represent the concept of human interest. These scales were used to obtain ratings on the standard concept of human interest in writing and on each written passage. The D statistic was computed between the two ratings on each passage and the standard. The more similar in meaning the rating to the standard, the more the human interest in a passage. Tannenbaum and Lynch (1960) working along similar lines developed a measure of sensationalism in writing, using the D statistic. A major advantage of the D statistic in this research is that each person serves as his own control, an important feature given the large variability in connotative meanings for the standards of human interest and sensationalism.

Conclusion

Given the ease of application and gathering of data with the semantic differential, researchers have often amassed large arrays of data without any concrete idea of how they are going to analyze it. Having gathered this mass of data some researchers feel compelled to analyze this data either on a scale-by-scale or dimensional basis. When it is discovered that the results for scales or for dimensions are difficult if not impossible to interpret in a meaningful way, some researchers will stretch for interpretations and be mistaken, or reject the semantic differential and abandon the large array of

data for more easily interpretable data. By building in the possibility of analysis with the D statistic at the outset of the research, it should be possible to get at interpretable results on meaningful variables with both efficiency and theoretical utility.

REFERENCES

Bushinski, R. Some Effects of Self-Directed and Teacher-Directed Instruction on Student Affect. Master Thesis. Northeastern University, 1969.

Lazowick, L.M. "On the Nature of Identification." J. of Abnormal and Social Psychology. 1955 51:pp175-83.

Lynch, M.D., "A Computer Projection Model for Estimating Election Returns." J. of Broadcasting Research. Summer 1967.

Lynch, M.D., et. al. "The Meaning of Human Interest: Four Dimensions of Judgment." Journalism Quarterly. 1967 44:pp673-8.

Lynch, M.D. and Lemkuhl, C. "Images of Public and Vocational School Reading Program in Beverly, Mass." Unpublished paper, Northeastern University 1971.

Lynch, M.D. "Connotative Images of Creative Adolescents." Unpublished paper, Northeastern University 1971.

Mahalanobis, P. C. "Historical Note on the D^2 Statistic." Sankhya 1948 9:pp237.

Mahalanobis, P. C., et. al. "Normalization of Statistical Variates and the Rise of Rectangular Covariates in the Theory of Sampling Distributions." Sankhya 1937 3:pp35.

Nugent, J. Some Correlates of Teacher Effectiveness. Master Thesis Northeastern University 1970.

Osgood, C. E., et. al. The Measurement of Meaning. Urbana, Ill.: University of Illinois Press 1957.

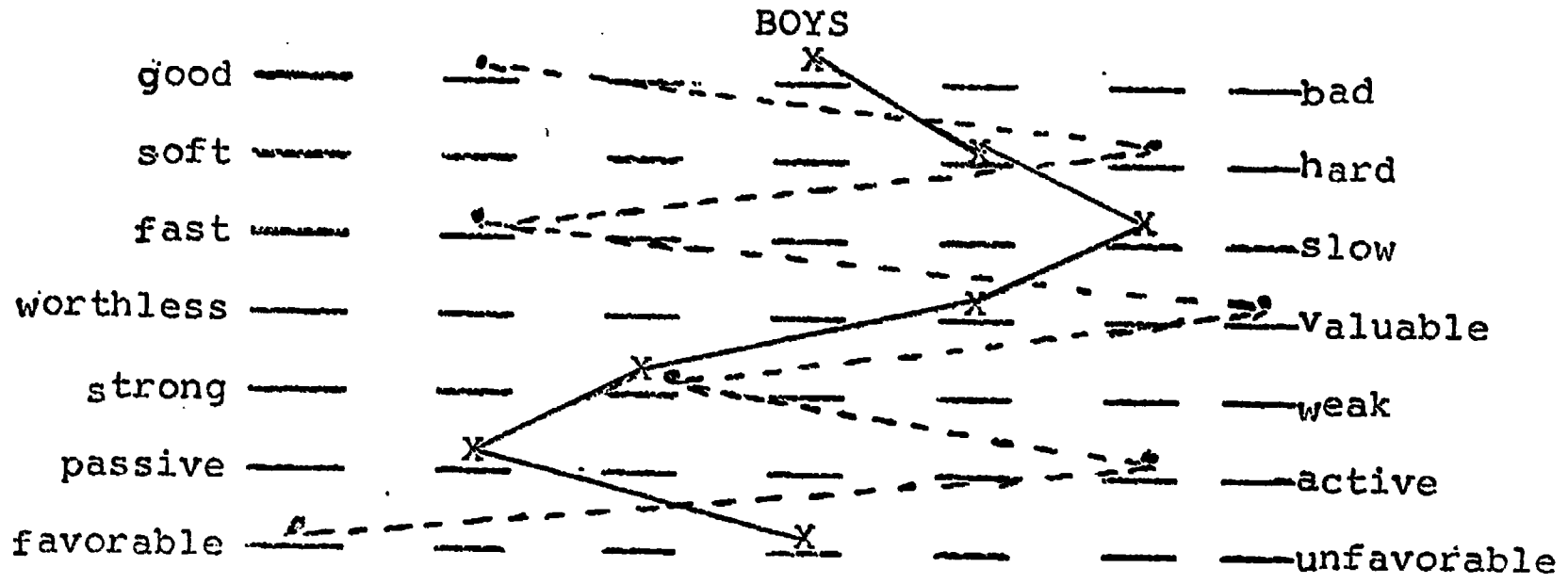
Osgood, C. E. and Snider, J. G. Semantic Differential Techniques. Chicago, Ill. Aldine 1969.

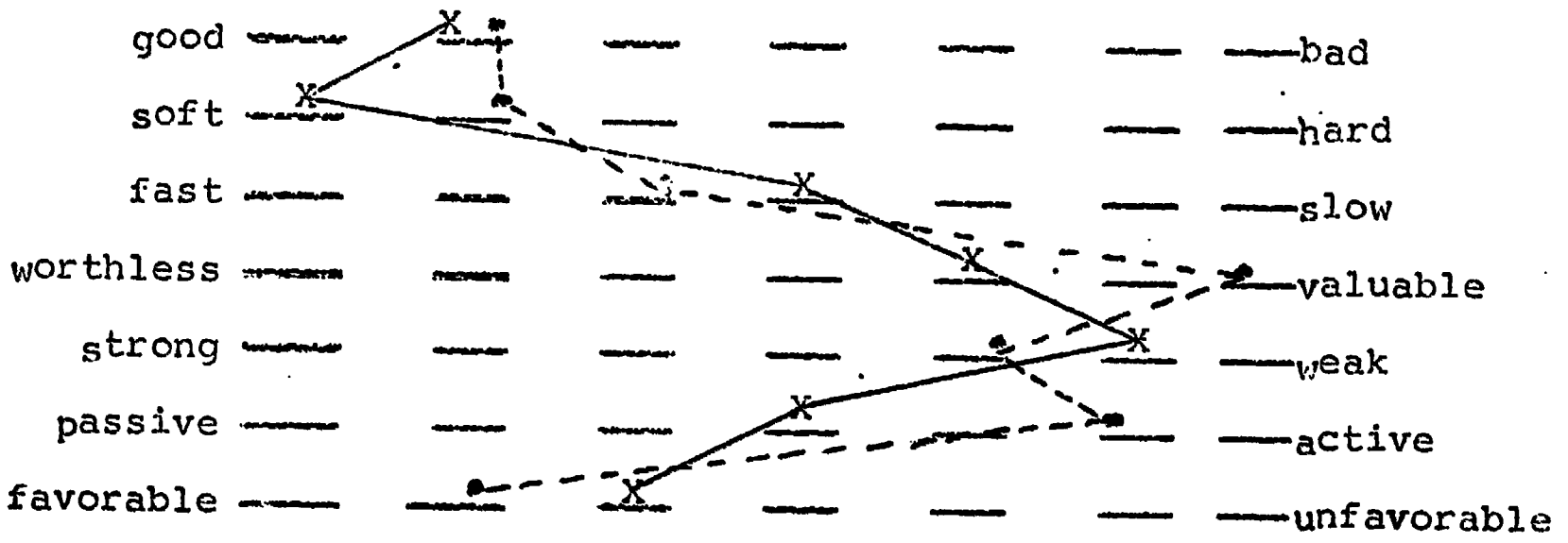
Tannenbaum, P. H. and Greenberg, B. The Great Debates: Influence on Candidate Images. University of Wisconsin, Center for Study in Mass Communications 1961.

Tannenbaum, P. H. and McLeod, J. M. "On the Measurement of Socialization." Public Opinion Quarterly. 1967 31:pp27-38.

FIGURE 1

PROFILE COMPARISONS BETWEEN SEMANTIC DIFFERENTIAL RATINGS ON MYSELF AND MY TEACHERS FOR 2 BOYS AND 2 GIRLS *





Myself as a student

My teachers

* (Illustrative purposes only)

TABLE I

SEMANTIC DIFFERENTIAL RATINGS ON EACH OF SEVEN SCALES FOR TWO CONCEPTS, MYSELF AND MY TEACHERS FOR TWO BOYS AND TWO GIRLS*

BOYS

	Myself 1	2	Total	Average	My Teachers 1	2	Total	Average
Good-Bad	5	4	9	4.5	6	7	13	6.5
Valuable-Worthless	5	5	10	5.0	7	7	14	7.0
Favorable-Unfavorable	4	5	9	4.5	7	7	14	7.0
Hard-Soft	5	6	11	5.5	6	7	13	6.5
Strong-Weak	4	6	10	5.0	5	6	11	5.5
Fast-Slow	3	2	5	2.5	6	6	12	6.0
Active-Passive	1	4	5	2.5	6	6	12	6.0

GIRLS

	Myself 1	2	Total	Average	My Teachers 1	2	Total	Average
Good-Bad	6	7	13	6.5	6	6	12	6.0
Valuable-Worthless	5	6	11	5.5	7	7	14	7.0
Favorable-Unfavorable	6	5	11	5.5	7	6	13	6.5
Hard-Soft	1	2	3	1.5	3	2	5	2.5
Strong-Weak	2	2	4	2.0	4	2	6	3.0
Fast-Slow	3	6	9	4.5	6	5	11	5.5
Active-Passive	5	4	9	4.5	6	6	12	6.0

* (Illustrative purposes only)

TABLE II

SEMANTIC DIFFERENTIAL RATINGS ON EACH OF THREE FACTORS, EVALUATION, POTENCY AND ACTIVITY FOR TWO CONCEPTS, MYSELF AND MY TEACHERS FOR TWO BOYS AND TWO GIRLS*

BOYS

	Myself		Total	Average	My Teachers		Total	Average
Evaluation								
Good-Bad	1	2			1	2		
Valuable-Worthless	5	4			6	7		
Favorable-Unfavorable	5	5			7	7		
	4	5			7	7		
	14	14	28	14.0	20	20	41	20.5
Potency								
Hard-Soft	5	6			6	7		
Strong-Weak	4	6			5	6		
	9	12	21	10.5	11	13	24	12.0
Activity								
Fast-Slow	3	2			6	6		
Active-Passive	1	4			6	6		
	4	6	10	5.0	12	12	24	12.0

GIRLS

	Myself		Total	Average	My Teachers		Total	Average
Evaluation								
Good-Bad	1	2			1	2		
Valuable-Worthless	6	7			6	6		
Favorable-Unfavorable	5	6			7	7		
	6	5			7	6		
	17	18	35	17.5	20	19	39	19.5
Potency								
Hard-Soft	1	2			3	2		
Strong-Weak	2	2			4	2		
	3	4	7	3.5	7	4	11	5.5
Activity								
Fast-Slow	3	6			6	5		
Active-Passive	5	4			6	6		
	8	10	18	9.0	12	11	23	11.5

* (Illustrative purposes only)

TABLE III

COMPARISONS OF SEMANTIC DIFFERENTIAL RATINGS BETWEEN BOYS AND GIRLS, AND CONCEPTS, MYSELF AND MY TEACHERS BY EACH OF THREE FACTORS, EVALUATION, POTENCY AND ACTIVITY*

Evaluation	Myself	My Teachers	Marginal
Boys	14.0	20.5	24.5
Girls	17.5	19.5	37.0
	31.5	40.0	71.5
Potency			
Boys	10.5	12.0	22.5
Girls	3.5	5.5	9.0
	14.0	17.5	31.5
Activity			
Boys	5.0	12.0	17.0
Girls	9.0	11.5	20.5
	14.0	23.5	37.5

* (Illustrative purposes only)

TABLE IV

D^2 COMPARISONS BETWEEN MYSELF AND MY TEACHERS FOR BOYS AND GIRLS COMPUTED ACCORDING TO THE FOLLOWING EQUATION $D^2 = \sum_{i=1}^{N} (X_{ij}-X_{ik})^2$ $D = \sqrt{D^2}$ *

i is a scale, j is a concept, k is a concept and n is the total number of scales

	Boys 1			2		
SCALES		$X_{ij}-X_{ik}$	$(X_{ij} X_{ik})^2$		$X_{ij} X_{ik}$	$(X_{ij}-X_{ik})^2$
Good-Bad	5-6=	-1	1	4-7=	-3	9
Valuable-Worthless	5-7=	-2	4	5-7=	-2	4
Favorable-Unfavorable	4-7=	-3	9	5-7=	-2	4
Hard-Soft	5-6=	-1	1	6-7=	-1	1
Strong-Weak	4-5=	-1	1	6-6=	0	0
Fast-Slow	3-6=	-3	9	2-6=	-4	16
Active-Passive	1-6=	-5	25	4-6=	-2	4
	D^2=		50			38
	$D = \sqrt{50}$ = 7.07			$\sqrt{38}$ = 6.16		

	GIRLS 1			2		
SCALES		$X_{ij}-X_{ik}$	$(X_{ij}-X_{ik})^2$		$X_{ij}-X_{ik}$	$(X_{ij}-X_{ik})^2$
Good-Bad	6-6=	0	0	7-7=	1	1
Valuable-Worthless	5-7=	-2	4	6-7=	-1	1
Favorable-Unfavorable	6-7=	-1	1	5-6=	-1	1
Hard-Soft	1-3=	-2	4	2-2=	0	0
Strong-Weak	2-4=	-2	4	2-2=	0	0
Fast-Slow	3-6=	-3	9	6-5=	1	1
Active-Passive	5-6=	-1	1	4-6=	-2	4
	D^2=		23			8
	$D = \sqrt{23}$ = 4.80			$\sqrt{8}$ = 2.83		

* (Illustrative purposes only)